- 1) If a subject is ready to make a bet at given odds and with a given stake, he will also be ready to make a bet on the same contingencies at the same odds but at any higher stake within his resources, or any lower non-negative stake.
- 2) If a subject is ready to make each of two bets on its own, he will also be ready to enter both bets simultaneously, provided this remains within his resources.
- 3) Sure-thing Principle for Preference: Let x be a variable. If, coteris paribus, for every possible value c of x the following statement is true:

If the subject knew that the value of x was c, he would prefer course of action A to course of action B.

then he will prefer course of action A to course of action E even in the absence of knowledge of the value of x.

- 4) Preference is transitive.
- 5) Admissibility: If two actions A and B differ only in that under a certain (possible) state of nature A yields more 'probability currency' than B, then A is preferred to B.
- 6) Direct Gift Principle: Two bets are equivalent if, under each member of an exhaustive set of pairwise incompatible states of nature, they have the same frequency probability of leaving the bettor in possession of the diamond.

(Two actions A and B are equivalent if for every action C, C is preferred to A iff C is preferred to B, and A is preferred to C iff B is preferred to C.)

7) (Limiting form of (3).) Let P be a proposition possibly true. If the following two statements are true:

If the subject knew that P were true, he would prefer course of action A to course of action B.

(The subject knows that) If P is false then courses of action A and B are identical in effect.

then he will prefer course of action A to course of action B in the absence of knowledge of the truth-value of P.

8) (Limiting form of (9).) Let P be a proposition. If the following two statements are true:

If the subject knew that P were true, he would have no definite preference between courses of action A and B.

(The subject knows that) If P is false then courses of action  $\Lambda$  and B are identical in effect.

then he will have no definite preference between courses of action A and B in the absence of knowledge of the truth-value of P.

9) Sure-thing Principle for Indifference: Let x be a variable. If, ceteris paribus, for every possible value c of x the following statement is true:

If the subject knew that the value of x was c, then he would have no preference either way between courses of action A and B.

then he will have no preference either way between A and B in the absence of knowledge of the value of  $\mathbf{x}$ .

## BAYESIAN CONDITIONALIZATION AND THE DUTCH BOOK THEOREM

By "Bayesian Conditionalization" I shall understand the inference by which a person, having accepted a probability <u>p</u> for a hypothesis <u>H</u> conditional on on the obtaining of certain evidence <u>E</u>, comes to accept <u>p</u> as the unconditional probability of <u>H</u> when the evidence <u>E</u> is, as it happens, actually obtained. Thus what I shall call "Bayesian Conditionalization" is what Hacking (1967) calls "the Dynamic Assumption".

By the "Dutch Book Theorem" I understand the theorem, formulated originally by de Finetti, that unless a man's personal probabilities satisfy the axioms of the probability calculus then it would be possible to place bets at his odds in such a way as to be sure of making money off him whatever actually happens. Shimony (1955) strengthened the conditions that a certain minimal degree of rationality seemed to impose on a man's personal probabilities by requiring that it should be impossible also to place bets at the man's odds in such a way as to be sure of not losing and at the same time to have a finite chance of winning. I shall call Shimony's argument as well as de Finetti's a "Dutch Book Argument" and aball also sometimes apply that description to an argument resting on the slightly weaker requirement of rationality that a man should not choose his odds in such a way as to allow the formulation of a betting policy against him which he himself must expect will win money off him on average.

definiing Bayesianism so as to include his Dynamic Assumption, Now Hacking (1967) (assures us (p. 316) that

neither the Dutch book argument, nor any other in the personalist arsenal of proofs of the probability axioms, entails the dynamic assumption. Not one entails Bayesianism. So the personalist requires the dynamic assumption in order to be Bayesian. It is true that in consistency a personalist could abandon the Bayesian model of learning from experience.

It is the purpose of the present note to argue that, contrary to the impression given by Hacking assertions, Dutch book arguments do go a considerable way towards justifying Bayesian conditionalization. Not only do such arguments rule out the obviously unwanted alternative coherence-preserving strategies

which seemed to be let in if one took the arguments as only capable of imposing a requirement of coherence (in the sense of consistency with the probability axioms) at any fixed time, but such arguments seem even to first sight to rule out certain kinds of deviation from strict conditionalization which on common sense grounds seem prima facie to be rational deviations.

The kind of deviation from conditionalization which Dutch book arguments can indeed rule out is well-illustrated in the following alleged objection to personalism (which he calls subjectivism) of Kyburg (1970):

The third objection, which applies only to the subjectivist theory, is already by the subjectivist's answer to the second objection: if you find the conclusion intolerable, change the premises. Since on the subjectivistic interpretation of probability I can adjust my probabilities in any way that I want to, provided only that they satisfy the rules of the calculus, there is nothing in principle that precludes my arranging my degrees of belief in such a way that I attribute a high degree of belief to the things that fit in with my preconceptions, and a low degree of belief to the things that don't fit in. That is: The relation of evidence to conclusion itself depends on the probability assignments that I make; I can in principle adjust these assignments so that the evidence supports (or does not badly undermine) the hypotheses that I want it to support, rather than those I don't want it to support. As Savage has observed, the probability calculus may show that we should modify some of our beliefs, but cannot show which of our beliefs should be modified.

Now it seems absurd that we should be able, so to speak, to change our beliefs retroactively after seeing to what further beliefs they lead. It is certainly poor science to decide which hypothesis to a ccept first and to evaluate the evidence concerning that hypothesis in the light of the hypothesis itself. The defense is that people just don't do that with their beliefs; people don't decide what to beliefe first and then calculate backwards, using the probability calculus, to find out what their prior beliefs should (should) have been. On the contrary, the way the probability calculus does generally function is to lead us to modify our beliefs in accordance with the evidence. Arather than to modify the impact of the evidence in accordance with our beliefs. From an abstract point of view this may seem like no more than a fortunate accident. From the point of view of the subjectivist it is simply a fact, a datum, and for the subjectivist that is sufficient. For the epistemologist or the scientist, however, this is the way the probability calculus ought to function, and indeed it is the whole point of having a probability calculus.

I claim that a Dutch book can be made against the kind of objectionable thus strategy that Kyburg describes and/that Kyburg's claim that the personalist that no a priori grounds for ruling such a strategy is incorrect. The objectionable strategy can be characterised in the following way. Instead of adopting his previous conditional probabilities as his new absolute was

once the evidence has come in, the theorist chooses to adopt higher absolute probabilities for his favoured hypotheses than those conditional probabilities and preserves coherence by revising his original priors upwards in retrospect. (in the Shimony sense) But to make a Dutch book/against such a theorist a need merely lay a certain sum in favour of the man's preferred hypotheses as a conditional bet with the condition being the obtaining of the unfavourable evidence. Then if that evidence is not obtained we lose nothing. If it is obtained 2 then place such a sum against the preferred hypotheses at the new odds as/recover me original stake if the preferred hypotheses eventually prove false. But since I know that the new odds are to be more favourable to the preferred hypotheses than the old conditional odds I will by this manner of placing bets have completed a book against the theorist which is such that I lose nothing if his theory proves false and I win money off him is his theory proves true. Thus suppose for example his old conditional odds were 3:1 against and his revised absolute odds were only 2:1 against and thus more favourable to his hypothesas than if he had carried out Bayesian conditionalization. Suppose I make a conditional bet of a \$1 in favour of his hypotheses and then after the evidence has come in (if it does come in) I make an absolute bet of against his hypotheses. If they then prove false I am quites if they then prove true I lose my \$2 but gain \$3 so am a \$ in hand. Of course in the situation envisaged of his new rates being specified I cannot know in advance/how much I shall gain off him if I do gain; however I know that I cannot lose and have a chance of gaining so long as (a) the hypotheses are such that an understanding can be obtained as to how such bets preferred can eventually be settled, (b) I know which his from the hypotheses are, (c) I know that he will invertence and the will be a superior and the will b if at all conditionalization/in the direction in which the odds/favour his preferred hypotheses, in comparison with the results of xxxxxxconditionalization. Andof course I can make money off such a man on average if I know only that on average he will depart from Bayesian conditionalization in a known direction. This direction need not be as in the above example. Thus not only can I also make a Dutch book against such a man by betting first against and then in

favour of the hypotheses which he is prejudiced against in the sense of failing to allow favourable evidence in relation to them to raise their probabilities successful as much as conditionalization would require, but I can also make a/book against in this sense a man who, while fiot prejudiced/for ar against any particular hypotheses, is nevertheless persistently overimpressed (or for that matter persistently underimpressed) by thereimental evidence, in the sense that he systematically departs from conditionalization by overadjusting (or for that matter underadjusting) his probabilities whenever really surprising experimental evidence turns up. Indeed the only systematic departure from Bayesian conditionalization against which it seems I cannot make a Dutch book, or at least be assured of winning money from the non-Bayesian on average, would be that of wholly unpredictable departures from conditionalization, which nevertheless agreed on average with the results of conditionalization.

These results are very encouraging for the Bayesian and seem fully to dispose of Kyburg's objections as well as to set Hacking's principal doubts at rest while leaving some freedom for manoevere by not making conditionalization absolutely mandatory on every occasion. However the worry now is whether these results are not indeed stronger than the Bayesian would really like.

In science, whenever an milkely areast occurs, it is common practice to re-examine the existing theories in the field in order to ascertain whether result the mark was really as unlikely according to those theories as it first appeared. This practice would reaise no difficulties for the Bayesian if eventuated it were the case either that the result of such re-examination marked as often in the experimental result finally appearing even less likely on the existing theories as it did in the finally appearing more likely than had originally been thought, or if such re-ex amination did indeed more often upwards than downwards but in such a way that this was true for the conditional probabilities on all (minally marked probabilities) competing hypotheses, and thus simply increased p(E) without altering

the conclusions of conditionalization. However there does seem some reason generally to believe that in reality neither of these situations/obtains and that does arise where experimental evidence taxabtainada which is quite anexpected on the basis of the best confirmed theory of the day, theorists do in fact succeed more often than can be attributed to chance in persuading their fellow scientists that the evidence was after all to be expected according to the previously well-confirmed theory, and that the arguments previously for the evidence, used for to indicate a low conditional probability/on the basis of that theory, had been defective.

Now such non-Bayesian behaviour on the part of scientists does not immediately lay them open to a Dutch book because they could perhaps evade it by the artifice of refusing to come to any agreement as to how bets on the scientific theories in question could in principle be settled. But this manoevre would not save them for long because obviously the whole argument could then be rephrased by replacing their theories by finite sets of their more striking unsettled specific predictions, where these predictions are made specific enough to allow bets on them to be just as settlable as any other bets. It does therefore seem that we can identify an aspect of the typical behaviour of scientists open which is such as to lay them to a Dutch book on average, namely one constructed by laying conditional bets aunfavour of their most successful theories where the condition is that of the occurrence of apparently highly unfavourable if these results occur bets results, and then/supplementing these/with unconditional bets against those theories once the scientists have "explained away" the unfavourable evidence. in certain respects The question now is, given that scientists laxxim do seem/to behave in such potentially a way as/to lay them selves open to such a Dutch book, is such behaviour really "irrational"?

The first point one needs to make in defence of the scientists is that such a revision of the original conditional probability of the evidence will typically bet through the discovery either of

original reasoning or of possibilities which had simply been overlooked and which on examination were clearly not at all unlikely. Both these cases amount essentially to the discovery of hidden incoherence in the original assignment of subjective probabilities. Once such incoherence is discovered it can hardly be irrational to put it right, especially since we can reasonably assume that those with whom the scientists wager read the learned journals and thus have access to such discoveries.

Of course the situation in which a person's original subjective probability distignments later turn out to be incoherent is not at all an unusual one, has been since as willishe obvious taxangx meraterx of Hacking's discussion. This does not however normally enable another person with no better information to make a Dutch book against the person; for if he has no better information he will be no better able to identify in advance the hidden flaws in the other man's reasoning. Also errors in reasoning will as often raise probabilities as lower them so there is no systematic way of capitalizing on the fact that there almost certainly are some errors somewhere. However the scientific example is peculiar in this respect in that the actual occurence or so it seems of the/very unlikely event is good evidence/for the presence of an error in reasoning in a particular direction and which will enable us to predict in confidence advance with some xxxxxxxxxxx the direction in which the posterior probabilities will in fact depart from the Bayesian ones. Of course we won't always be right, sometimes there will be no change, occasionally there will be a can expect to change in the other direction, but it does seem that we will be right often enough to make a steady profit. Very unlikely events do happen occasionally and from time to time highly successful theories are indeed overturned but we also know that more often than not a re-examination of the apparently unfavourable evidence leads to what everyone will agree is a full vindication of the previously well-established theories.

The problem to this. Can a scientist rationally say that such and

such an experimental result in exceedingly likely unlikely given the that
well-supported theory to which he subscribes, but/should such a result
be obtained he would conclude that it was probably not unlikely after
all, given his theory, but that there must almost certainly have been
some oversight in his calculations? One wants to say both that it is certainly rational to say and believe this, and that this tantamount
to offering to let anyone make an almost certain Dutch book at one's own
expense. This seems to me a paradox.

At first one might hope to be able to block the paradox by disallowing the relevant conditional bet. Clearly, one might say, it is irrational for such a scientist to allow anyone to place such a conditional bet in these circumstances. But it seems that this way out will not do because one can for any conditional bet devise an equivalent combination of unconditional bets with the same effect: thus instead of betting on H given E the bettor can bet on H.E and then cover himself against the possibility of -E by making an appropriate bet against E, i.e. one which will leave him quits if E doesn't occur. The resulting bet system is equivalent to a bet on H given E. And there seems no reason to say it is irrational for the scientist to accept bets for or against E or for artagainst H.E.

A second attempt at a way out might be to say that the original conditional C probabilities of E is dependent not just on H but also on the hypothesis/that certain calculations are correct, and that if this dependence is made explicit all will be well, for the occurrence of E will then be seen to be clear evidence C. against/thereachesis/approximations/area thereachesis/approximations/area thereachesis/approximations/area thereachesis/approximations/area thereachesis are incorrect then we have not reason prior to the occurrence of E to suppose that they have led to an error in one direction rather than in the other, and therefore no reason to take p(E, H.C) smaller than attraction p(E, H.-C). Indeed had we had such a reason we ought already to have allowed for this and adjusted our original p(E, H) accordingly.

Formally the situation is this:  $p(E, H.C) =_{def} p(E, H)$ . But also we must have p(E, H) = p(E, H.C)p(C) + p(E, H.C)p(C), with p(C) = 1 - p(C). But then, by some simple algebra, either p(C) = 1 and we are back where we started, or p(E, H.C) = p(E, H.C). But it now follows that, since the truth or falsity of C cannot effect p(E, H.C) = p(E, C) = p(E, C). Hence if we are to be good Bayesians we must even reject the claim that the occurrence of E can be taken to bonstitute evidence against C.

It seems to me that we are left with only one way out of the apparent paradox and that this is simply to reject all non-Bayesian intuitions here. If a scientist claims both that a certain experimental result has a very claims low probability on the basis of a certain successful theory, and that if would the that experimental result actually occurred it is more likely than not that it would be probability on the basis of the theory after all, then such a scientist is being irrational and the continuous of its being irrational and the continuous of its being irrational and the continuous of science appears to show that evidence that is at first sight highly unfavourable to otherwise successful theories proves more often than not perfectly in accordance to be acknowledged to be acknowledged as a certain successful them on closer examination, then either we have been witnessing a rare unlikely sequence of a certain successful theories must have been fooling themselves and everybody else,

The history of coelestial mechanics is here an interesting test case for it seems to be riddled with examples of observational results which were fix first believed to conflict with the Newtonian theory but which turned out when the calculations were redone more carefully fully to accord with that theory. Is this not a clear case where a Dutch book could be made by betting in favour of the Newtonian theory while the evidence seemed against it and then betting that theory against it once the evidence was reconciled with it satisfactorilly? (Again I simplify the betting situation somewhat: the bets would have to be on the

Newtonian theory qua good approximation for all ordinary phenomena of I have in mind here especially the public history of the publicar of the degree of oblickness of the earth, of the motion of the moon's nodes, of the secular exceleration of the moon, and of the laugheder of Unenes.

(on pethaps ments also good approximation for those puticular phenomena) coelestial mechanics and this phrase will have to be construed in such a way the outcome of that potential evidence makes/the bet decidable in such awa way that one can say that the issue has now been decided (in favour of the Newtonian theory, qua such an approximation), but hadn't been decided in the eighteenth or early nineteenth century. The precise details of the way this is done do explicating not matter, since the problem is not that of/my original phoney bets on "the Newtonian theory", but merely one of showing that there could be realistic wagers which would show precisely the features to which I am drawing attention.) Closer examination of the historical situation in astronomy seems to show that the situation was in fact not one in which a Dutch book could in principle have been made. The condition that must be satisfied in order that a Dutch book not be makeable is that for any hypothesis E, p(K, E) at one time must be equal to the expected value of  $p(\underline{H})$  at any time after the observation of  $\underline{E}$ has occurred. The expectations can be taken either as those of the scientist or of his appoint the man who wishes to wager with him: in the former case we have a Dutch book in the sense that the scientist can expect and with no better information than himself to be able better to wager with him so as to toxxiandxnoxahancexofilosinxxani win money off him on average; in the second case we have a Dutch book in the sense that the scientists opponent can place bets so as to expect to be able to make money of the scientist. Since no man's expectations here are intrinsically better than any others both kinds of situation seem equally objectionable. Now at first the astronomy situation seems to be one where the evidence starts highly unfavourable to the Newtonian theory (i.e. it disagrees with it's predictions) but where there is a good chance that more careful computations will enable us to show later that it satisfactorily ascords with that theory. This would certainly seem a situation where a Dutch book could be made. But/closer analysis it becomes clear that even though initially the evidence disagrees with the predictions it does not follow that it is hence

highly unfavourable to the Newtonian theory. This for the MENNER following conditional reason: the/probability of the actual evidence on the basis of the negation of the Newtonian theory is not all that higher than it is on the basis of the Newtonian theory: in fact if the chance of the calcuaations being correct is less than fifty per cent, the former probability will be less than twice the latter probability, but both will be very small because there are almost uncountable ways in which one can depart from the Newtonian predictions and none of which is particularly favoured by the mere denial of the Newtonian theory (historically it is not a matter of pyedente which while disagreeing more-or-less with Newton agreed with some/plausible rival theory: that would be a different situation). So when we calculate p(H, E) to begin with we get something which we can assume is already pretty close to 1 (p(H)) times, say, one in a million, divided by itself plus something pretty close to 0 (p(-H)) times, say, two in ammillion, and/we still end up with something pretty close to 1. And even if one supposes a time early on in the history, when the Newtonian theory has had fewer successes, and, say, the initial odds are only two to one in its favour, directly disconfirmed one still finds on the basis of the above assumptions that the motivamentalize predi**tti**on still existence only brings the odds down to evens. So what seemed prima facie to be initiallyhighly unfavourable evidence turns out on closer examination to be only very mildly unfavourable evidence. Furthermore, once the Newtonian theory has acquired a/high subjective probability, even the brilliant triumph of a Laplace in finally reconciling the Newtonian observed quantitative predictions with the/secular acceleration of the moon according to the ancient observations of eclipses, cannot raise that probability much in percentage terms making already (which is what counts viz-a-viz/a Dutch book) because it will/be so a hundred per cent anyway. (We must also allow of course for the possibility that even a Laplace's calculations prove incorrect, as indeed happened in more than precisely this case as Adams showed/half a century later. The residual discrepancy thus uncovered again was serious, and not properly quantitatively accounted for for/another century: it an effect of the frictional drag of the tides in the earths shallower seas, so they say.) Finally, and most

to reconcile the actual observations with the Newtonian theory does not merely after leave the probability of the Newtonian theory the same as it was when the unfavourable evidence first came in, but actually substantially lowers it.

When we do some model calculations taking into account all these facts we find both that we get a satisfactory qualitative account of the relevant features of the history/astromomy and that our model does not allow a Dutch book to be made. Thus to take an oversimplified example suppose that the eventually subjective probability for the Newtonian theory/accounting satisfactorily subjective for the evidence in question is .9, suppose that the/probability of any particular calculation containing a mistake is 2, and suppose for simplicity, that agreement between the calculation and the evidence merely by chance subjective has a primary probability of O. (Obviously we would not want to assume the latter in a more realistic calculation) Now it turns out that on these subjective probability assumptions there is a warrance 45% that the first calculation will theory and observation and if it does the new probability for the theory (as construed in this example) will become 100% (xix tradity by definition): subjective there is a/probability of 55% that the first calculation will fail to demonstrate agreement and in this case bhaditionalization would give a new probability just maker over 8 for the theory of just under 82%. But 55 times/22xis equal to 45 times 10 so this result is also just what is needed to prevent a Dutch book then were being made. Of course if we carried out an indefinite number of recalculations subjective probability the anamor of eventually getting agreement between the theory and the observations would be .9. However this will only raise the dubjective probability of the theory from .9 to 1, whereas the alternative eventuality of failing indefinitely to get agreement has anxinitial subjective probability of .l and would reduce the subjective probability of the theory to 0. So again the expectation value of the change in the subjective probability is zero and no Butch book

can be made. Such calculations in themselves of course just confirm the internal consistency of probability theory. But their point is to show in some detail how it is that the fact that we have ggod reason to believe, in a domain like coelestial mechanics, that evidence unfavourable to the Newtonian theory will nearly always prove to be explainable away satisfactorily, does not show, as it firfirst appears to, that the scientists are failing to change their subjective probabilities in accordance with Bayesian conditionalization, or that area if they are failing to do so they are failing in such a that area to depart from conditionalization on average in area a way area lays themselves open to a Dutch book.

I conclude that Dutch book theorems can be used to justify Bayesian conditionalization as a temporal process, as well as the coherence of a persons subjective probability assignments at any particular time, in the sense that alternative systematic policies for retaining coherence at each particular time but departing from conditionalization seem to be ruled out. At first sight that conclusion seems not only to rule out the kind of prejudiced behaviour which Kyburg seemed to think Bayesians could rule out some only by fiat, but seems also to rule out/seemingly rational behaviour of scientists which prima facie does seem to depart from Bayesian conditionalization in a way which would invite the making of a Dutch book. However on closer examination this prima facie impression proves to be mistaken and bo both the scientists and the Bayesians are vindicated.

Jon Dorling Chelsea College London.

Refs.

I. Hacking: Slightly More Realistic Personal Probability, Phil. Sci. 34 311-325. (1967)

H.E. Kyburg: Probability and Inductive Logic (Macmillan, 1970), pp. 73-74.

A. Shimony: Coherence and the Axioms of Confirmation, J. Symb. Log. 20 1-28. (1955)